

GABRIELE ROMANATO

Menu

Sulla validazione dei documenti HTML5

In questo articolo tratteremo quello che effettivamente può diventare un problema qualora volessimo validare il nostro documento in HTML5.

HTML5, HTML 4 e XHTML

HTML5 ha rimosso alcuni vincoli imposti da **XHTML** e ha introdotto delle novità, ossia:

- Gli elementi vuoti possono anche non avere lo slash finale prima della chiusura del tag.
- Gli attributi con valore booleano possono essere scritti con la vecchia notazione HTML 4, ossia:

```
<input type="text" name="test" required />
```

- Alcuni elementi deprecati nella DTD Strict di XHTML sono ora disponibili in HTML5, come ad esempio `iframe`
- Non c'è più la case-sensitivity per i valori di attributi e nomi di elementi, ossia si possono indistintamente usare le maiuscole e le minuscole.
- Sparisce il riferimento alla DTD nel DOCTYPE.
- Un documento HTML5 valido può anche non includere l'elemento `body`. Questa libertà è data dal fatto che i browser di norma "aggiustano" la struttura DOM della pagina qualora questa manchi di alcuni elementi strutturali.
- Alcuni attributi diventano opzionali. Ad esempio gli elementi `style` e `script` possono anche non avere l'attributo `type`.
- HTML5 può essere servito come `application/xhtml+xml` (come XHTML), ma in quel caso l'elemento radice deve avere il namespace XHTML.

- In HTML5 alcuni elementi, come `acronym`, sono stati dichiarati obsoleti.
- HTML5 introduce nuovi elementi pur mantenendo sostanzialmente la retrocompatibilità con XHTML e HTML 4.

Il problema dei siti dinamici

Fintanto che si creano documenti **HTML5** statici, la validazione di HTML5 non presenta grandi problemi. Ma quando, come in WordPress o in altri CMS, la marcatura viene generata dinamicamente, noi non abbiamo il totale controllo di quello che sarà il risultato finale.

In WordPress, ad esempio, validare il nostro tema non crea problemi, perchè siamo noi a decidere la struttura HTML generata dai template tags. Ma quando inseriamo widget o plugin, il risultato potrebbe invalidare la nostra marcatura perchè in genere questi prodotti di terze parti possono usare dei tag non compatibili con HTML5 o che comunque non rispettano le regole sintattiche di base di HTML.

Per esempio un bottone di Twitter inserito tramite un elemento `iframe` può creare problemi se gli URL non sono codificati o se le entità `&` non sono scritte come `&`.

Quello che possiamo fare in questo caso è analizzare il codice HTML generato da widget e plugin e modificarlo tenendo presente le regole di HTML5.